主动测量技术对非结构对等网络的影响

沙瀛

摘要：主动测量技术已经成为研究对等网络特征属性和用户行为的一种重要基础性手段。但是主动测量技术对对等网络的影响还是未知的，目前还缺乏这方面的详细分析和形式化研究。针对上述问题，本文首先介绍了当前对等网络测量的发展和现状，分析了当前所面临的主要挑战性问题，着重介绍了我们在主动测量技术对非结构对等网络的影响方面研究工作的初步进展。我们首先提出了主动测量网络模型，着重研究了主动测量技术对对等网络度分布的影响。研究发现随着主动测量网络规模的增加将导致度分布的相变现象。理论分析和仿真实验证明一个小规模的高“度”的主动测量网络要优于大规模的低“度”的主动测量网络。

关键词：对等网络；主动测量；度分布

1　引言

准确描述对等网络（peer-to-peer, P2P）拓扑结构的动态行为，并分析其网络性能，无论是对于网络运营商还是对等网络协议开发者来说都具有重要的参考价值。而要完成这些目标，一个重要的基础性工作就是精确的网络测量。

当前面向对等网络主要有主动和被动两种网络测量方式。由于被动测量仅适用于测量流经监测点的对等网络流量，无法将监测范围覆盖到整个对等网络，而且该方法只能适用于流量性能测量，无法获取网络搜索、路由、应用层传输等信息。因此主动测量技术有其不可替代的作用。粗略地说，主动测量技术一般使用网络爬虫(crawler)，将其主动加入对等网络，获取相关的网络特性和节点属性，比如“节点（Peer）”的 IP 地址、端口号以及所有可以通过对等网络协议获取的元数据(metadata)信息。使用主动测量技术，我们可以获得三类对等网络行为信息，即：(1). 微观行为特性，比如对等网络的拓扑结构、延迟、内容可用性、上传/下载比；(2). 拓扑动态行为特征，比如稳定核心覆盖(stable core overlay，即顶级拓扑中在线时间超过特定门限值的节点集合)连通性的变化、“节点”会话时间和下载时间的分布、以及会话时间的关联性等；(3). 扰动行为(churn)，即用户加入/退出对等网络系统所引起的动态变化。

终上所述，主动测量技术已经成为研究对等网络的拓扑、度分布等属性特征和用户行为特征的一种重要基础性手段，有着不可替代的作用。但是主动测量技术由于其本质特征，必然对原有对等网络的拓扑、流量、节点的策略选择以及用户行为决策等产生影响，目前还缺乏这方面的详细分析和形式化研究。因此，如何定量地评测主动测量技术的精确度，进而设计出主动测量的校准算法以还原出待测量的对等网络的更准确全面的信息，是具有重要研究意义的课题。

我们主要针对主动测量技术对对等网络的影响开展了研究，具体思路是：首先对主动测量节点和主动测量网络建模；对其对对等网络的各方面影响进行理论分析，即对主动测量网络与对等网络叠加后构成的混合网络进行研究，获得形式化结果；提出主动测量校准算法，以期通过校准来获得对等网络更为真实准确的信息；最后通过仿真实验和实际对等网络环境验证模型和算法的有效性。最终目标是通过对主动测量技术对对等网络影响及其校准技术的

研究，建立主动测量技术的统一评价标准，使建立在主动测量基础上的对等网络相关研究更加真实可靠，并进一步指导对等网络的设计和优化。目前我们在对等网络的度分布方面已获得了初步结果，此项研究工作已获得国家自然科学基金的支持（国家自然科学基金面上项目，编号：61070184），并于 2011 年启动。

2 对等网络主动测量技术现状

本节首先介绍当前对等网络测量的研究热点，然后介绍对等网络测量领域仍然存在的挑战性问题。

2.1 国内外研究现状及发展动态

当前围绕对等网络测量的研究主要侧重如下几个方面：

(1) 测量数据获取方法与理论建模

此类研究工作较多，主要关注测量方法的设计（如网络爬行器）、对测量数据的分析、对理论模型的仿真和测量验证，如文献[17]、[21]、[30]、[31]、[32]等。萨若优（S. Saroiu）等人[17]率先使用主动网络爬虫对当时最为流行的两种对等网络系统 Napster[1]和 Gnutella[2]进行了测量研究。梁建（音译，J. Liang）等人[18, 19]对 KaZaA[3]系统的网络行为进行了深入的探索与分析，设计出专门的网络爬虫，首次对 KaZaA 系统中的文件污染状况[20]进行了测量与分析。但是上述研究工作都缺乏对测量的准确性等相关问题的理论分析。只是在数据分析中提到网络爬行器的设计对结果的影响：由于对等网络的动态变化，如果为了获得更为准确的网络快照，爬行器爬行时间要短，但爬行时间过短则限制了爬行的范围。斯图泽巴赫（D. Stutzbach）和瑞加亚（R. Rejaie）[28]研究了 Gnutella 网络爬虫的快照准确性问题，但采用的手段主要是实际测量爬行器不同频率、速度、爬行时间对对等网络拓扑的影响，缺乏理论分析和数学模型。伊斯达（T. Isda）等人[23]指出主动测量会发送正确协议中不期望的数据包，在大规模情况下可能会引起麻烦，比如造成入侵检测系统的报警。文中指出了测量行为可能成为异常行为，会给系统造成影响，但是没有解决对系统影响的问题。

由此可以看出虽然对对等网络已经进行了大量的测量，但是对于测量过程的有效性、测量结果的准确性还缺乏普遍的关注。而这些都影响了通过测量数据对对等网络进行分析的准确性和真实性。

(2) 大规模对等网络中的取样问题

由于对等网络系统的大规模、动态性，直接获取全局行为数据是不现实的，因此取样是感知这些系统的有效方法。取样的关键性问题在于如何获取代表性的数据，如何由较少取样数据导出全局的性质。

斯图泽巴赫和瑞加亚[1]主要关注如何获取节点属性(如节点的度、链接带宽、共享文件数等)的代表性节点，指出对对等网络系统的取样会引入两方面误差：（一）时间（由于节点不同的生命周期）、（二）拓扑属性（节点“度”的差异）。取样更倾向于高连接度的节点。依照现有的研究[2,5]，许多节点生命期是很短的，而有些节点则会存

1 一种在线音乐服务
2 一种文件共享网络
3 一款点对点文件共享工具

在很长的时间，随着取样时间的增加，取样集合会增加大量的短生命期的节点。

萨若优等人[4]采用广度优先或深度优先的方式对节点取样，这两种方式都会对短生命周期的节点产生偏向。另外一种方式是随机游走，其中一种简单的方案是随机游走长度 r，将终端节点作为取样点，再行走 r，依次类推。虽然此种方法对某些类型的图效果较好，但是其效率较低。按照图论，建议 $r \geq \log|V|$ [6][7]。

若将对等网络映射为图，则有些问题可以借鉴图论中的研究结果。博洛巴什（B. Bollobás）[8]、杰若姆（M. Jerrum）[9]定义了一类共享某些属性（如度分布等）的图，然后采用特定的随机算法产生所有此类的图。库珀（C. Cooper） 等人[10]使用这种方法证明他们的覆盖（overlay）重建算法可以产生良好属性的图。克里希纳穆尔蒂（V. Krishnamurthy）[11]、斯塔姆普夫（M.P.H. Stumpf）[13]等人从一个大的图中通过取样产生出一个保持了原始图属性的代表性的子图。

斯图泽巴赫和瑞加亚[24]主要关注的问题是：如何从一个未知的大规模的动态变化的图中获取无偏的代表性节点。莱斯科维茨（J. Leskovec ）等人[25]研究了图的进化，集中关注稠化作用(网络变得密集)、直径缩小（随着网络的增长，直径减少）等属性，并用新的图生成器来解释这些属性。他们的观察以年为跨度。

取样技术影响了主动测量节点的选择。但是目前还缺乏对对等网络主动测量误差的定量理论分析。如果主动测量技术对对等网络产生了扭曲，则即使采用了无偏的取样技术，获得的信息也是不准确的。因此需要在首先解决主动测量技术对对等网络的影响基础上，解决测量的误差和校准问题。

(3)　**搜索覆盖率的统计分析**

特普斯特拉（W.W. Terpstra）等人[26][27]基于随机多图提出了一个基于概率的穷尽搜索系统发泡风暴（BubbleStorm），发泡风暴将整个系统分为多个查询“泡”（Query Bubble）和数据“泡”（Data Bubble），若两者有交叉，则可以搜索到相关的内容。此系统利用了节点间带宽不同的特性。察沃斯(Y. Chawathe)[14]、甘特斯蒂斯(C. Gkantsidis) [15] 、吕(音译， Lv) [16] 等人用随机游走作为非结构对等网络搜索的基础，但是搜索只需要在行走过程中确定一个特定数据的位置，而不关心某些节点更倾向被选择，这一点不同于无偏取样。

(4)　**建立验证基于测量的网络研究有效性的标准**

奥尔曼(M. Allman) [29]提出了反馈测量(reactive measurement)，即将测量看作一个过程(process)而不是一个现象(event)，要由前期测量的结果来判断是否需要，如需要应选取何种下一步的测量，这说明测量是一个动态的过程，也从一个侧面说明了需要测量的一个客观评价标准；克里希纳穆尔蒂等人于 2008 年[22]提出需要建立验证基于测量的网络研究的有效性标准，并对基于测量的网络研究提出了验证标准的初步想法。他们共提出了 4 个问题：（一）测量的质量、（二）测量的统计分析是否合适、（三）建模方法的科学价值、（四）验证的完全性。实际上对等网络的测量也有同样的问题，针对对等网络的测量也同样迫切需要建立统一的评价标准。为了建立对等网络测量的统一评价标准，首先就需要对主动测量对对等网络的影响进行充分的研究。

因此主动测量技术对对等网络影响的研究是关系到测量准确性、有效性的关键性基础问题，需要在对主动测量技术的本质特征充分分析总结的基础上，建立主动测量的数学模型，从理论上分析其对对等网络产生的影响，获得形式化的结果；在此基础上提出主动测量的校

准算法以获得待测对等网络更为准确全面的关键特征和信息。

2.2 当前研究面临的问题

当前已有大量的研究人员对各种对等网络系统进行了充分的测量和分析，得出了很多有益的结果。但是当前主动测量技术还面临下面的问题：

(1) **主动测量技术对对等网络的影响未知**

对等网络系统的规模越来越大，主动测量系统的规模也越来越大，但是这里有一个问题，就是主动测量技术对对等网络的影响是未知的。

对等网络的主动测量技术通常要定制对等网络系统的客户端，形成网络爬虫，主动加入对等网络，获取相关的网络特性和节点属性。为了获取相关的信息，网络爬虫通常遵守对等网络的基本协议，但是为了测量的效率和准确性，又会对协议进行一定的修改。当网络爬虫达到一定的规模，必然会对对等网络的特性、性能等指标产生影响，影响对等网络中相关节点、服务器的决策，进一步会影响用户的决策。对等网络原始节点的策略也会受到插入对等网络的主动测量节点影响。虽然每个测量节点发起的连接都是清楚的，但是对等网络的原始节点由于这些测量节点的增加而导致的策略等方面的变化则是未知的。

由于缺乏主动测量技术对对等网络的影响的研究，也就无从判断测量过程的有效性、测量数据的准确性。但是目前还没有发现相关的形式化分析和理论研究。

(2) **现有取样技术无法保证获得的信息是正确的**

如前所述，取样是感知这些系统的有效方法，由取样获得的局部信息推导出全局信息。随着对等网络系统规模的增加，其取样的规模也在增加。取样技术影响了主动测量节点的选择。但是目前还缺乏对对等网络主动测量误差的定量理论分析。如果主动测量技术对对等网络产生了扭曲，则即使采用了无偏的取样技术，获得的信息也是不正确的。因此需要首先在了解主动测量技术对对等网络的影响的基础上，解决测量的误差和校准问题，这样采用合理的取样技术才是有意义的。

(3) **缺乏主动测量技术的客观评价标准**

随着对等网络测量工作的深入，有关测量的有效性验证标准等问题正逐步得到人们的关注。虽然已经有了大量的测量工作，但是对于测量过程的有效性、测量结果的正确性等问题还缺乏客观的评价标准。没有客观的评价标准，则无从评价测量方法的优劣，也不利于测量数据的共享，无从评价测量数据的质量，也就不可能在数据处理阶段采用合理的技术手段以避免或克服测量数据的先天缺陷，最终也就不能保证通过测量对理论模型进行验证的准确性完整性。

综上所述，当前主动测量的主要问题是：主动测量技术对对等网络的影响是未知的；现有主动测量技术不能获得全局的网络快照；主动测量的准确度还缺乏定量的理论分析，主动测量作为研究对等网络的重要手段还缺乏客观的评价标准。

3 我们的工作

本节首先介绍我们的研究路线，然后介绍我们在主动测量技术对对等网络度分布的影响方面所获得的初步结果。

3.1 基本研究路线

我们首先需要研究主动测量节点的行为模型和由主动测量节点构成的主动测量网络的行为模型。观测到的测量结果实际上是待测对等网络与主动测量网络的叠加结果。在前两个研究成果基础上，将测量结果进行合理的拆分，可获得待测对等网络更真实准确的网络特性和节点属性等信息，也就可以建立主动测量的校准算法。

具体研究包括以下几个方面：

- 主动测量节点的研究：包括主动测量节点的属性模型、行为模型等，重点在于其与正常对等网络节点的比较；
- 主动测量网络的研究：包括主动测量网络模型、其本质特征、与待测的对等网络的异同等；
- 主动测量对对等网络的影响：分析对等网络受主动测量影响的属性和影响的方式、建立主动测量网络与对等网络的叠加模型及其形式化结果；
- 主动测量校准算法及其验证：将测量结果进行合理拆分，获得待测对等网络的真实准确网络特性等信息，并在仿真和实际网络环境中验证模型和算法的有效性。

3.2 主动测量技术对对等网络度分布的影响

度分布是刻画对等网络的一个重要特性，因此我们首先研究主动测量技术对非结构对等网络的度分布的影响，以此作为本课题研究的起点。首先定义节点 $\boldsymbol{i}$ 的度 $\boldsymbol{k_i}$ 为与该节点连接的其它节点的数目。网络中所有节点 $\boldsymbol{i}$ 的度 $\boldsymbol{k_i}$ 的平均值称为网络的平均度，记为 $\langle \boldsymbol{k} \rangle$。网络中节点的度的分布情况可用分布函数 $\boldsymbol{p(k)}$ 来描述。 $\boldsymbol{p(k)}$ 为网络中度为 $\boldsymbol{k}$ 的节点占所有节点数的比例，也就是随机选取的一个节点的度为 $\boldsymbol{k}$ 的概率。

大量研究表明许多实际网络的度分布可以用幂律形式 $\boldsymbol{p(k) \sim k^{-r}}$ 来描述。如果一个函数 $\boldsymbol{f(x)}$ 具有如下性质：对任意给定常数 $\boldsymbol{a}$，存在常数 $\boldsymbol{b}$ 使得 $\boldsymbol{f(ax)=bf(x)}$，则称该函数是“无标度”的。

幂函数是具有这种性质的唯一函数。这表明具有幂律度分布的网络没有明显的特征标度，此类网络称为无标度网络。在一个大规模的无标度网络中，绝大部分的节点的度相对很低，但存在少量的度的数值相对很高的节点。因此，无标度网络是非均匀的，而那些少量的度相对很高的节点称为网络的“中心节点”。

对对等网络的主动测量的整体行为可以看作是待测量的对等网络与测量网络的叠加：

$$\boldsymbol{G_m = G_0 + G_a}$$

其中：$\boldsymbol{G_m}$ 为测量所得的网络，即待测对等网络叠加了主动测量网络后形成的混合网络；

$\boldsymbol{G_0}$ 为待测的对等网络；

$\boldsymbol{G_a}$ 为主动测量网络。

3.2.1 待测对等网络 G_0 的生成模型

许多研究表明非结构对等网络符合无标度特性，其度分布可以用 $\boldsymbol{p(k) \sim ak^{-r}}$ 来描述。因此我们可以采用巴拉巴斯-阿尔伯特模型（Barabasi-Albert model，BA model，即 BA 模型）[33]来产生 $\boldsymbol{G_0}$ 的模型。首先是实际网络的增长(Growth)特性，即网络是开放的，网络规模并非

固定而是不断增大的。其次，实际网络具有优先连接(Preferential Attachment)特性，即新的节点更倾向于与那些具有较高连接度的“大”的节点相连接。基于网络的增长和优先连接特性，BA 无标度网络模型的构造算法如下[33]

(1) 网络规模增长：从一个具有 m_0 个节点的网络开始, 每次引入一个新的节点, 并将该节点连接到 m 个已经存在的节点上。

(2) 节点优先连接：一个新节点与一个已经存在的节点相连接的概率 Π_i 与节点 i 的度 k_i 之间满足以下关系：

$$\Pi_i = \frac{k_i}{\sum_j k_j}$$

在经过 t 步后，这种算法产生一个有 $N = t + m_0$ 个节点，m_t 条边的网络。当 N 充分大时，网络的度分布函数可由指数为-3 的幂律函数描述，即网络中连接度为 k 的节点的概率与 k^{-3} 成正比，且网络的平均度为 $2m_0$。

当 $t \to \infty$ 时，得到不依赖于时间的幂律度分布：

$$P_0(k) = 2m_0^2 k^{-3}$$

3.2.2 混合网络 G_m 的生成模型

首先分析主动测量网络 G_0 在对对等网络度分布的影响方面所具有的特性：

- 为了提高测量获得的网络快照的准确性，主动测量的时间不能持续太长；
- 因为主动测量的时间不能持续太长，所以主动测量节点的生命周期也不会持续太长，但是在主动测量时间周期内，新增的节点应该主要是主动测量节点；
- 由于对等网络动态化、大规模特性，主动测量节点只占整个对等网络中的一小部分；
- 主动测量节点的度一般都高于正常对等网络节点的度，只有这样测量节点才可以获得尽可能多的信息，但在测量周期内，新增连接应主要是测量节点与已存在的节点的连接；
- 主动测量节点通常只能获得当前对等网络局部信息，并向其中部分节点发起连接。

考虑到主动测量网络 G_a 的特性，所提出的混合网络 G_m 的生成模型，主要是模拟主动测量的过程。设已知待测对等网络共有 N 个节点，M 个连接，主动测量获得的本地对等网络规模为 γ，则可通过重复下面的步骤建立混合网络 G_m 的生成模型：

■ **步骤 1：建立本地对等网络。**随机或采用某种取样算法从整个对等网络中选取 γ 个节点，这 γ 个节点对主动测量节点来说构成了一个本地对等网络，记为 Δ；

■ **步骤 2：增加节点。**以概率 P 在本地对等网络 Δ 中增加一个有 m_1 个连接的新节点。任何一个本地网络中已存在的节点 i 在时刻 t 与之发生连接的概率 $\Pi_\Delta[k_i(t)]$ 正比于节点 i 的平均度，即：

$$\Pi_\Delta\left[k_i(t)\right] = \frac{\gamma}{N + pt} \times \frac{k_i(t)}{\sum_{j\in\Delta} k_j(t)};$$

■ **步骤 3：增加连接。**以概率 q 在本地网络中已存在节点之间增加 m_1 个新连接。新连接一端是随机选择的，另一端是以概率 $\Pi_\Delta\left[k_i(t)\right]$ 选择的；

■ **步骤 4：移除连接。**以概率 $1 - p - q$ 在本地网络中移除 m_2 个连接，这些连接都是在

本地网络中随机选择的。

这里，$0 < p < 1$, $0 \le q < 1$， $m_1 < \gamma < N + pt$, $m_2 \le m_1$ 。

步骤 1 可以采用多种方法实现由全局对等网络到本地局部网络的映射; 步骤 2 对应于增加一个新的测量节点，因为在主动测量期间，新增加的节点以测量节点为主，因此这个模型里主要考虑测量节点; 步骤 3 对应于增加测量节点和正常对等网络节点之间的连接，这些连接包括测量节点与正常节点之间的连接和正常节点之间的连接。由于节点性能的限制，系统在单位时间内只能增加有限的连接，因此我们设置上述两个操作的参数都是 m_1 。最后一步模拟了对等网络的动态特性，即单位时间内会有一定的测量节点和正常节点之间断开连接。

当然这里的对等网络模型和主动测量网络与对等网络的叠加模型都是相对比较简单的，随着研究的深入会逐步采用更复杂的模型以接近实际系统。

3.2.3 混合网络 G_m 的度分布

本节利用 Barabasi_Albert(巴拉巴斯-阿尔伯特)连续域(mean field)理论分析混合网络与待测网络在度分布方面的差异。由上节可知，在每个时间步有 $m_1 \le \gamma \le N + pt$ 。于是，我们可以分别考虑两个极端情况：

(1) $\gamma = N + pt$

在这种情况下，本地对等网络就是整个对等网络网络，其度分布为：

$$P(k) = \frac{p^2 m_1}{(p+q)} \langle k_0 \rangle k^{-3} \propto a k^{-b}$$

此处 $\langle k_0 \rangle$ 是待测对等网络“度”的平均值。

这说明在这种情况下混合网络仍然是一个无标度网络，即如果主动测量的规模较小，并且主动测量网络可以获得整个对等网络的信息。主动测量只影响系数 a ，而不影响幂指数 b 。

(2) $\gamma = m_1$

此时的度分布为：

$$P(k) = \frac{-1}{\alpha m_1} e^{\frac{p}{\alpha}} e^{\frac{-pk}{\alpha m_1}} \propto e^{\frac{-p}{\alpha m_1}k} , \quad \alpha = p + \frac{2q}{m_1 - 1}$$

此时混合网络已符合指数分布。

从上述两个极端的例子我们可以看出，如果 $\gamma \approx m_1$,则 $P_m(k)$ 接近指数分布；如果 $\gamma \approx N + pt$ ，则 $P_m(k)$ 接近无标度分布；当 $m_1 < \gamma < (N + pt)$，则 $P_m(k)$ 从指数分布过渡到无标度分布。如果本地对等网络规模足够大，主动测量节点主要与已存在的正常对等网络节点相连，混合网络仍然保持无标度网络，则主动测量网络对待测对等网络影响较小；但是如果本地对等网络规模较小，则混合网络将过渡到指数分布，这意味着主动测量网络对待测对等网络影响较大。因此主动测量网络应该获得尽可能多的待测对等网络节点的信息。

3.2.4 仿真实验

我们主要采用 MATLAB 仿真了待测网络 G_0 和混合网络 G_m 的生成模型。采用如下标识：

$P_0(k)$ 和 $P_m(k)$ 分别用 P_0， P_m 表示； G_0 表示为 $G_0(N_0,m_0)$，其中 N_0 是 G_0 网络节点总数，m_0 是每个时间步增加的连接数；G_m 表示为 $G_m(N_m,m_1,m_2,p,q,G_0)$，其中 N_m 是 G_m 网络节点总数，m_1 是每个时间步增加的连接数，m_2 是每个时间步移除的连接数， p 是增加节点的概率，q 是增加连接的概率，G_0 是初始待测对等网络。我们仿真 G_0 的规模从 2000 到 20000。因为结果相似，下面只用 $N_0=5000$ 为例说明。

(1) P_0 和 P_m 的比较

G_0(5000,5)和 G_m(5010, 10, 1, 0.6, 0.3, G_0)的度分布如图 1 所示，X 轴表示节点的度，Y 轴表示度分布（双对数坐标）。由图 1 可知其度分布接近一条直线，所以两个网络的度分布都符合无标度分布，两者没有明显的差异，说明小规模主动测量网络对对等网络的影响较少。

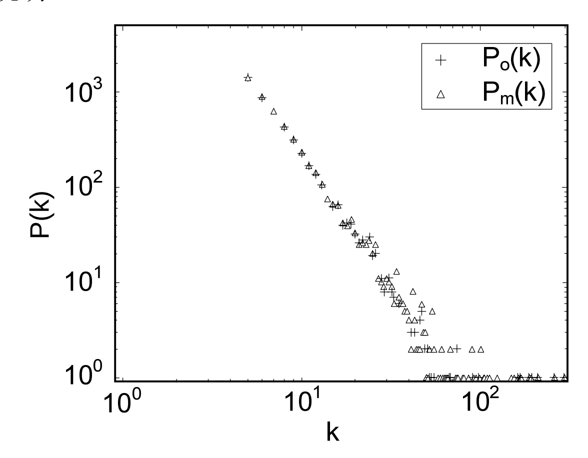


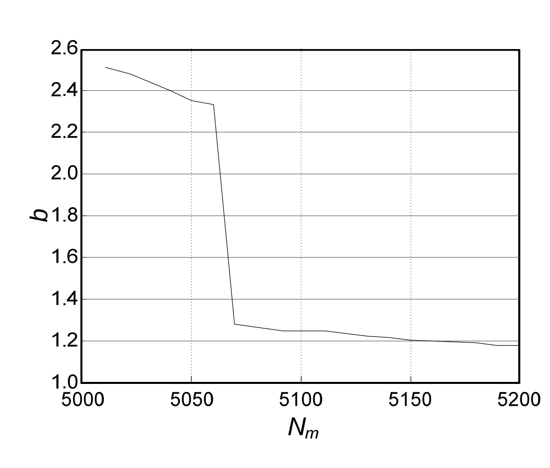
图1. 待测网络度分布 P_0 和混合网络度分布 P_m

(2) N_m 对 P_m 的影响

随着 N_m 的增加，导致系数 a 和幂指数 b 急剧下降，而产生相变现象（图 2），此时 $N_m=5070$。如图 3 所示，当 N_m 超过 5070 的时候，其度分布已经严重偏离了无标度分布。我们观察到即使很小规模的主动测量网络（1%左右）就会对度分布产生严重的影响（表 1）。因此在设计主动测量网络的时候，应该采用小规模高“度”的测量节点而不是大规模的低“度”的测量节点。

表 1 当发生相变时 N_0 和 N_m 值

N_0	N_m	%	N_0	N_m	%
3000	3030	1.0	7000	7080	1.1
4000	4050	1.3	8000	8090	1.1
5000	5070	1.4	9000	9100	1.1
6000	6060	1.0	10000	101000	1.0



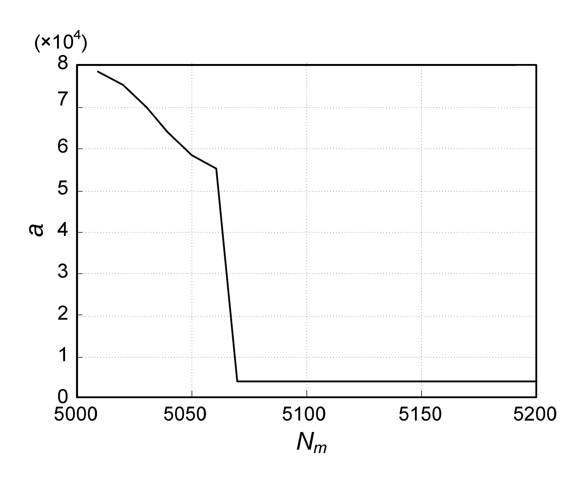


图2. N_m 对系数 a 和幂指数 b 的影响，X 轴表示 N_m 从 5000 到 5200。.Y 轴分别是 a 和 b。

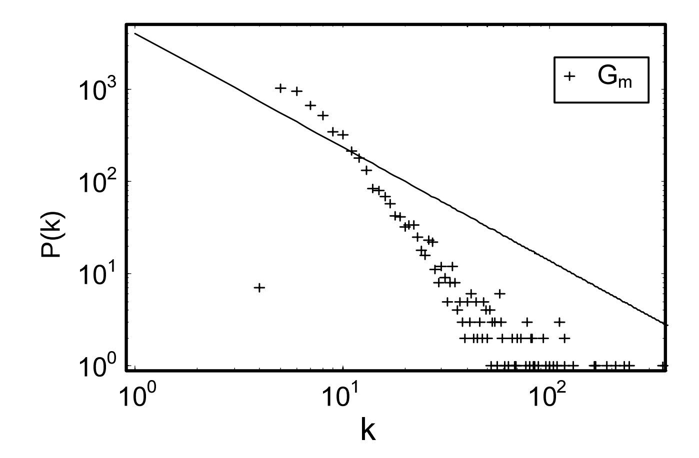


图3. P_m：G_0(5000, 5)；G_m(5130,10,1,0.6,0.3,G_0)

4 总结与未来工作

本文首先介绍了对等网络测量的发展和现状，指出了当前对等网络测量的主要问题是主动测量技术对对等网络的影响是未知的，由此影响了对对等网络测量结果的判断和校正，不利于建立统一的测量评价标准。然后介绍了我们在主动测量技术对对等网络影响方面的工作：提出了基本的研究路线；对主动测量技术对对等网络度分布的影响获得了初步结果。下一步工作将进一步研究主动测量技术对对等网络其他方面的特性的影响，并进而提出主动测量的校正算法，以获得正确真实的测量结果，提出统一的测量评价标准。

参考文献：

[1] S. Daniel, *R. Reza, Sampling Techniques for Large, Dynamic Graphs*, in Global Internet Symposium,http://mirage.cs.uoregon.edu/pub/sampling-gi06.pdf

[2] M. Izal, G. Urvoy-Keller, E. W. *Biersack, et al. Dissecting BitTorrent: Five Months in a Torrent's Lifetime*. In PAM, 2004.

[3] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips. *The Bittorrent P2P File-sharing System: Measurements and Analysis. In International Workshop on Peer-to-Peer Systems (IPTPS), 2005.*

[4] S. Saroiu, P. K. Gummadi, and S. D. Gribble. *Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts.* Multimedia Systems Journal, 9(2), 2003.

[5] D. Stutzbach, R. Rejaie, and S. Sen. *Characterizing Unstructured Overlay Topologies in Modern P2P File-Sharing Systems*. In Internet Measurement Conference, 2005.

[6] C. Gkantsidis, M. Mihail, and A. Saberi. *Random Walks in Peer-to-Peer Networks*. In INFOCOM, 2004.

[7] L. Lov´asz. *Random walks on graphs: A survey*. Combinatorics: Paul Erd¨os is Eighty, 2, 1993.

[8] B. Bollobás. *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs.* European Journal of Combinatorics, 1, 1980.

[9] M. Jerrum and A. *Sinclair. Fast uniform generation of regular graphs*. Theoretical Computer Science, 73, 1990.

[10] C. Cooper, M. Dyer, and C. Greenhill. *Sampling regular graphs and a peer-to-peer network*. In Symposium on Discrete Algorithms, 2005

[11] V. Krishnamurthy, M. Faloutsos, M. Chrobak, L. Lao, J.-H. Cui, and A. G. *Percus. Reducing Large Internet Topologies for Faster Simulations*. In IFIP Networking, 2005.

[12] V. Krishnamurthy, J. Sun, M. Faloutsos, and S. Tauro. *Sampling Internet Topologies: How Small Can We Go?* In International Conference on Internet Computing, 2003.

[13] M. P. H. Stumpf, C. Wiuf, and R. M. May. *Subnets of scale-free networks are not scale-free: Sampling properties of networks.* Proceedings of the National Academy of Sciences, 102(12), 2005.

[14] Y. Chawathe, S. Ratnasamy, and L. Breslau. *Making Gnutella-like P2P Systems Scalable*. In SIGCOMM, 2003.

[15] C. Gkantsidis, M. Mihail, and A. Saberi. *Random Walks in Peer-to-Peer Networks*. In INFOCOM, 2004

[16] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. *Search and Replication in Unstructured Peer-to-Peer Networks.* In International Conference on Supercomputing, 2002.

[17] S. Saroiu, PK. Gummadi, SD. *Gribble, A measurement study of peer-to-peer file sharing systems*. In: Proc. of the Multimedia Computing and Networking (MMCN) 2002. January 2002.

[18] J. Liang, R. Kumar, KW. Ross, The KaZaA overlay: *A measurement study.In:Proc.of the 19th IEEE Annual Comp Communications* Workshop.2004.http://cis.poly.edu/~ross/papers/KazaaOverlay.pdf

[19] J. Liang, R. Kumar, KW. Ross. *Understanding KaZaA*.2004. http://cis.poly.edu/~ross/papers/UnderstandingKaZaA.pdf

[20] J. Liang, R. Kumar, KW. Ross.*Pollution in file sharing systems*.In:Proc.of the IEEE Infocom 2005.2005.ht cis.poly.edu/~ross/papers/pollution.pdf

[21] Xiaoming Wang; Zhongmei Yao; *D. Loguinov.Residual-Based Measurement of Peer and Link Lifetimes in Gnutella Networks*; INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE 6-12 May 2007 Page(s):391 - 399.

[22] B. Krishnamurthy, W. Willinger *What are our standards for validation of measurement-based networking research?*, HotMetrics 2008 June 6, 2008 Annapolis, MD First Workshop on Hot Topics in Measurement & Modeling of Computer Systems

[23] T. Isda, M. Piatek, A. Krishnamurthy, T. E. Anderson: *Leveraging BitTorrent for End Host Measurements, Passive and Active Measurment*, 8th International Conference, PAM 2007.

[24] D.Stutzbach, R. Rejaie, *On Unbiased Sampling for Unstructured Peer-to-Peer Networks, Internet Measurement Conference(IMC)*, Proceedings of 6th ACM SIGCOMM conference on Internet measurement. 2006.

[25] J. Leskovec, J. Kleinberg, and C. Faloutsos. *Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations*. In KDD, 2005.

[26] W. W. Terpstra, J.Kangasharju, C. Leng, A. P. *Buchmann, BubbleStorm: Resilient, Probabilistic, and Exhaustive Peer-to-peer search*, Proc. ACM SIGCOMM 2007.

[27] W. W. Terpstra, C. Leng, and A. P. Buchmann. *BubbleStorm: Analysis of Probabilistic Exhaustive Search in a Heterogeneous Peer-to-Peer System*. Technical Report TUD-CS-2007-2, Technische Universit¨at Darmstadt, Germany, 2007.

[28] D. Stutzbach, R. Rejaie, *Evaluating the accuracy of captured snapshots by peer-to-peer crawlers*, Passive and Active Measurement Workship, Extended Abstract, PAM 2005.

[29] M. Allman, V. Paxson, *A Reactive Measurement Framework*, Passive and Active Network Measurement (2008), pp. 92-101.

[30] Mo Zhou, Yafei Dai, and Xiaoming Li, *"A Measurement Study of the Structured Overlay Network in P2P File-Sharing Systems,"* International Journal of Advances in Multimedia, vol. 2007, Article ID 27958, 9 pages, 2007

[31] Zhao Yang, Hou Xiaoxiao, Yang Mao, Dai Yafei. *Measurement Study and Application of Social Network in the Maze P2P File-Sharing System*. In proceeding of International Workshop on

Peer-to-Peer Information Management (P2PIM), May 2006, Hong Kong.

[32] Bin Cheng, Xiuzheng Liu, Zheng Zhang, Hai Jin, Lex Stein, and Xiaofei Liao, *Evaluation and Optimization of a Peer-to-Peer Video-on-Demand System, Journal of Systems Architecture*, Vol.54, No.7, Elsevier Science, July 2008, pp.651-663.

[33] D. J. Watts, S.H. *Strogatz. Collective Dynamics of 'Small-world' Networks.* Nature,393(6684):440-442,1998.

作者简介：
沙 瀛： 中国科学院计算技术研究所信息安全研究中心助理研究员 shaying@ict.ac.cn

（ 上接第 48 页 ）

[8] C. Allauzen, M. Crochemore and M. Raffinot, "Efficient Experimental String Matching by Weak Factor Recognition", in *Proc. 12th Annu. Symp. on Combinatorial Pattern Matching*, Jerusalem, July 1–4, 2001, pp. 51-72.

[9] Snort Rule: http://www.snort.org/snort-rules

[10] L. Salmela, J. Tarhio, J. KytÄojoki. Multi-pattern string matching with q-grams. ACM Journal of Experimental Algorithmics (JEA), 11, 2006. http://www.jea.acm.org

[11] 杨毅夫．2009．面向深度包检测的正则表达式匹配技术．北京：中国科学院计算技术研究所研究所．37-39

[12] Sidhu R, Prasanna V K. 2001. Fast Regular Expression Matching using FPGAs. the 9th Annual IEEE Symposium on Field-Programmable Custom Computing Machines. Rohnert Park：227-238

作者简介：
何慧敏： 中国科学院计算技术研究所信息安全研究中心，硕士生
hehuimin@software.ict.ac.cn
刘 夏： 中国科学院计算技术研究所信息安全研究中心，硕士生
姜 磊： 中国科学院计算技术研究所信息安全研究中心，博士生